STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

# Nancy Grace Roman Space Telescope Science Operations Center (SOC) Data Management System

Harry Ferguson

November 15, 2021

# SOC Responsibilities & Design

## STScI provides the Science Operations Center (SOC) within a distributed Ground System Architecture

**Planning & Scheduling**
All mission observations

**WFI Data processing**
Details depend on mode

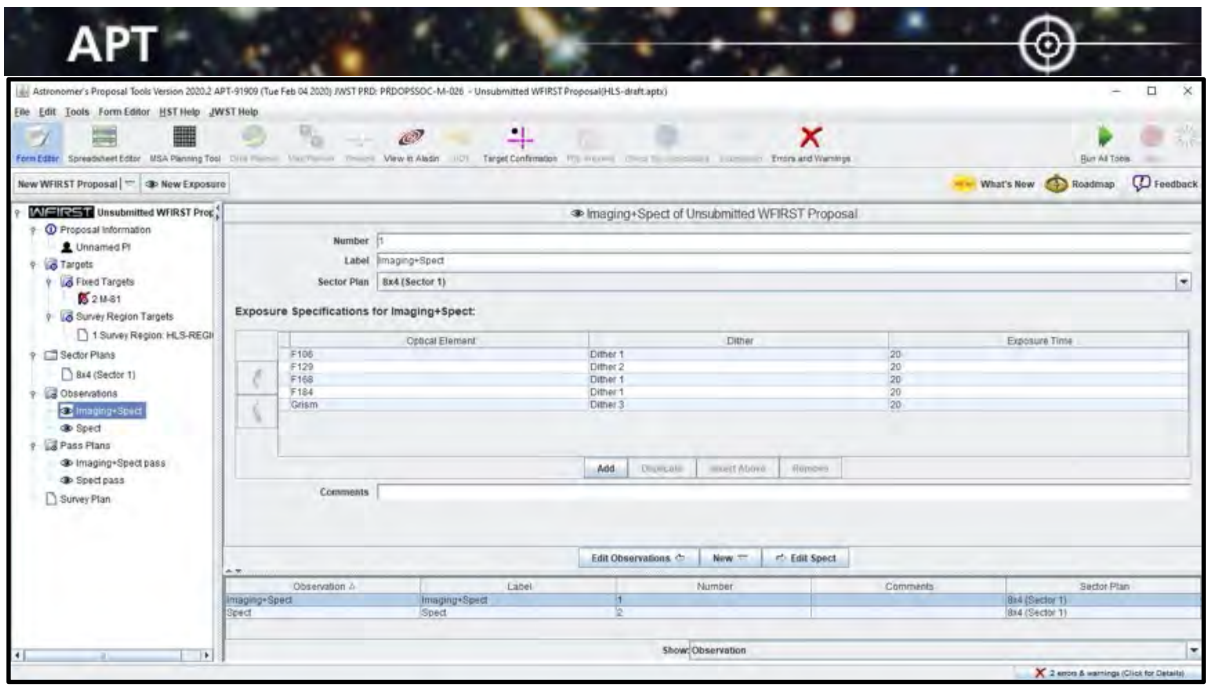
**Archive**
for all observations

**Community interface for WFI imaging**
User support , documentation, public outreach

## Remaining Ground System functions: IPAC, GSFC, …

# Planning and Scheduling System

- Adapted from HST/JWST “Astronomer’s Proposal Tool”
- Significant heritage for other parts of scheduling system
- Campaign-scheduling approach to long-term planning
  - Optimize surveys while allowing interleaving of different programs




Roman APT Example: High Latitude Wide Area Survey with imaging+spectroscopy

# Working Groups, User Support, Documentation, etc.

- **SOC Scientists perform a range of activities**
  - (co-)chair project-wide Working Groups with project and community scientists on a range of scientific and technical topics
  - Organize Hack Days, Focus Meetings, Jamborees, etc.
  - User support of Science Teams (or are themselves co-investigators)
  - White papers and documentation to inform the community about Roman's scientific capabilities & opportunities

**Solar system science with the Wide-Field Infrared Survey Telescope**

Bryan J. Holler et al.

**Astrometry with the Wide-Field Infrared Space Telescope**

The WFIRST Astrometry Working Group:

**Roman coronagraphic operations: lessons learned from the Hubble Space Telescope and the James Webb Space Telescope**

John H. Debes et al.

Etc, etc, etc……

Journal of Astronomical Telescopes, Instruments, and Systems

**An Ultra Deep Field survey with WFIRST**

Anton M. Koekemoer (STScI), R. J. Foley (UCSC), D. N. Spergel (Princeton/CCA), M. Bagley (UT Austin)R. Bezanson (Pittsburgh), F. B. Bianco (NYU), R. Bouwens (Leiden), L. Bradley (STScI), G. Brammer (NBI), P. Capak (Caltech), I. Davidzon (Caltech), G. De Rosa (STScI), M. E. Dickinson (NOAO), O. Doré (JPL), J. S. Dunlop (ROE), R. S. Ellis (UCL), X. Fan (Arizona), G. G. Fazio (CfA), H.

**ASTRO 2020**

arXiv.org > astro-ph > arXiv:1907.07184

Search... Help | Advanced

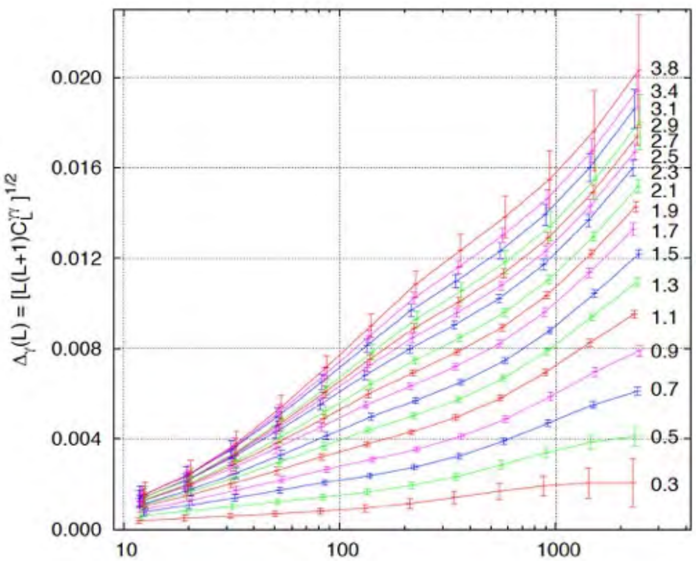
Astrophysics > Instrumentation and Methods for Astrophysics

**On the need for synthetic data and robust data simulators in the 2020s**

Molly S. Peeples (STScI/JHU), Bjorn Emonts (NRAO), Mark Kyprianou (STScI), Matthew T. Penny (Ohio State), Gregory F. Snyder (STScI), Christopher C. Stark (STScI), Michael Troxel (Duke), Neil T. Zimmerman (GSFC), John ZuHone (Harvard-Smithsonian CfA)

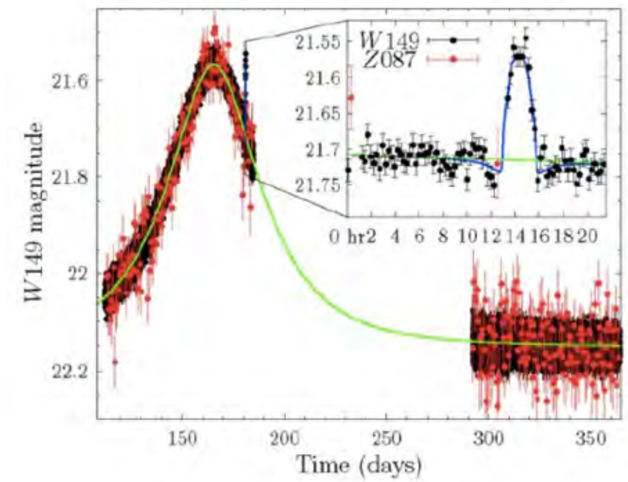
# WFI/Imaging
# Calibration Planning

**0.05% PSF shape (impacts cosmic shear)**



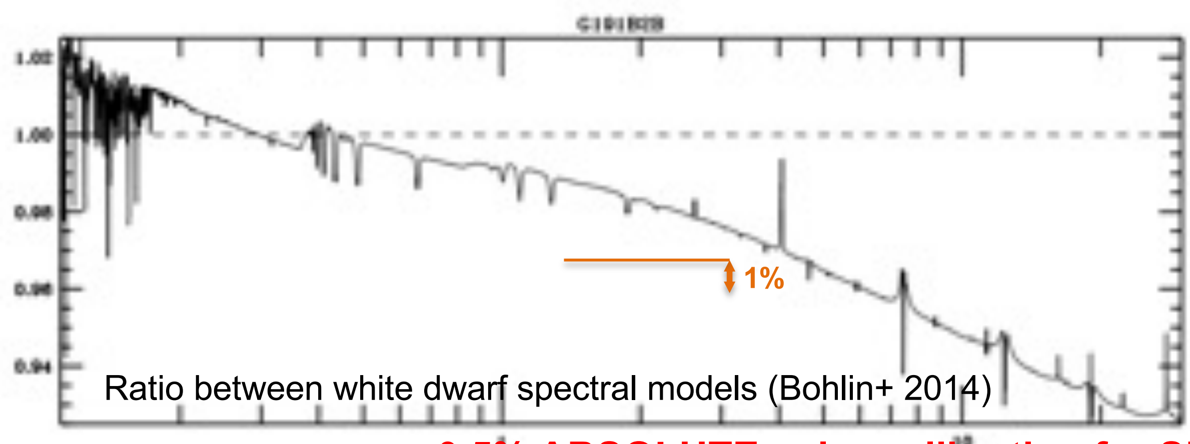

Expected shear power spectrum[L]
(Science Definition Team report, Spergel+ 2015)

**0.1% Photometric stability for microlensing (maps to planet mass uncertainty)**




Simulated microlensing event (S. Carey)

- Required accuracies up to ~10x better than what is available on other missions (e.g., Hubble)




Ratio between white dwarf spectral models (Bohlin+ 2014)

**0.5% ABSOLUTE color calibration for SNe (maps to luminosity distance vs redshift)**

STScI | SPACE TELESCOPE SCIENCE INSTITUTE

# The Data Management System

## NASA Astrophysics *Big Data*:

- Data accumulated per week likely to be >>100x *Hubble*
- Both catalogs and pixel-level data sets provide unique science opportunities
- Downloading and processing exceeds resources typically available
- Most research will be *archival*, given the survey nature of the mission

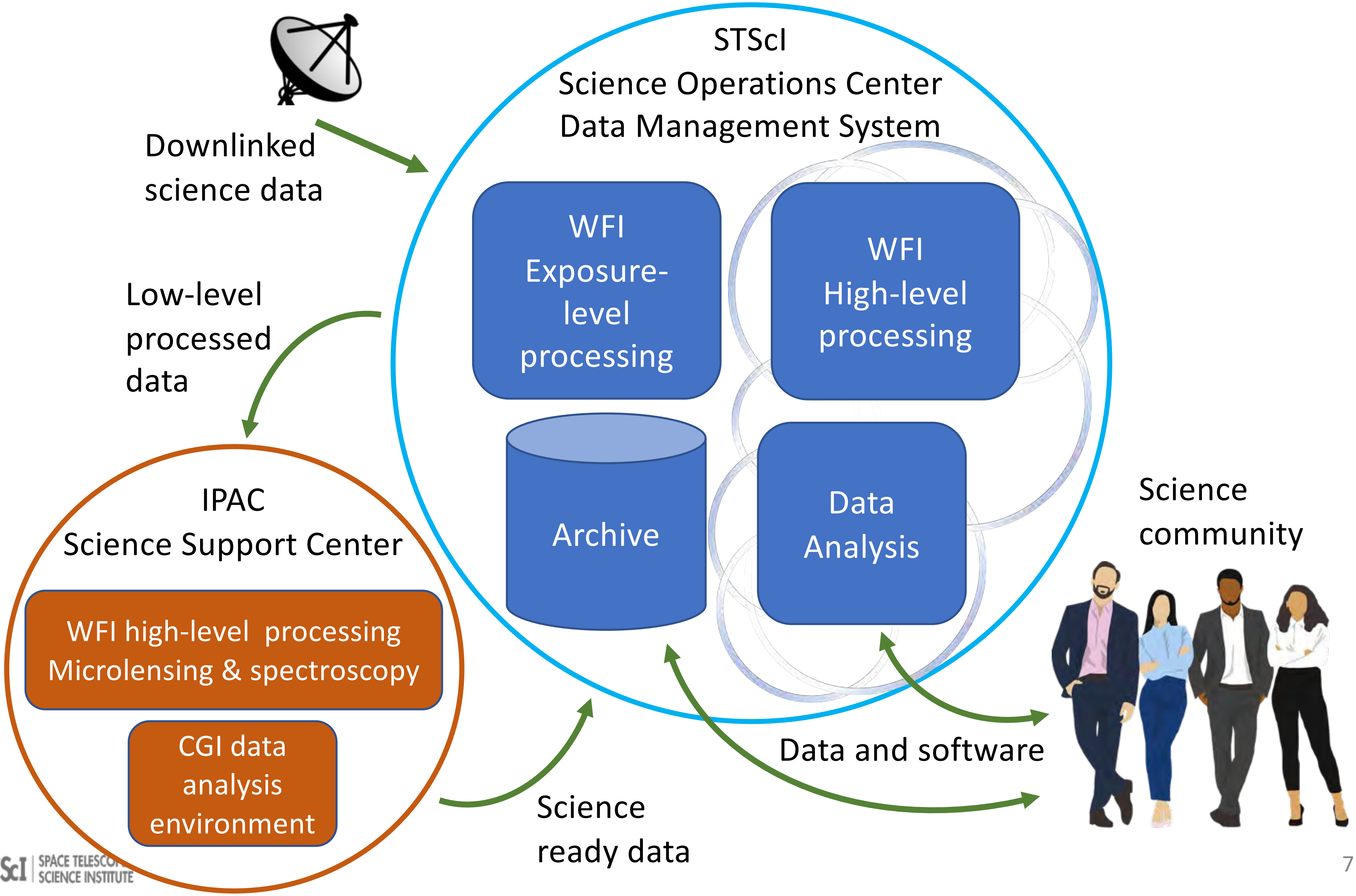
## Science data products from multiple mission partners

- Calibrated and mosaiced images, extracted spectra, catalogs, etc.
- Staged in the cloud and co-located with significant computational resources
- Open source, modular imaging pipeline facilitates custom reprocessing

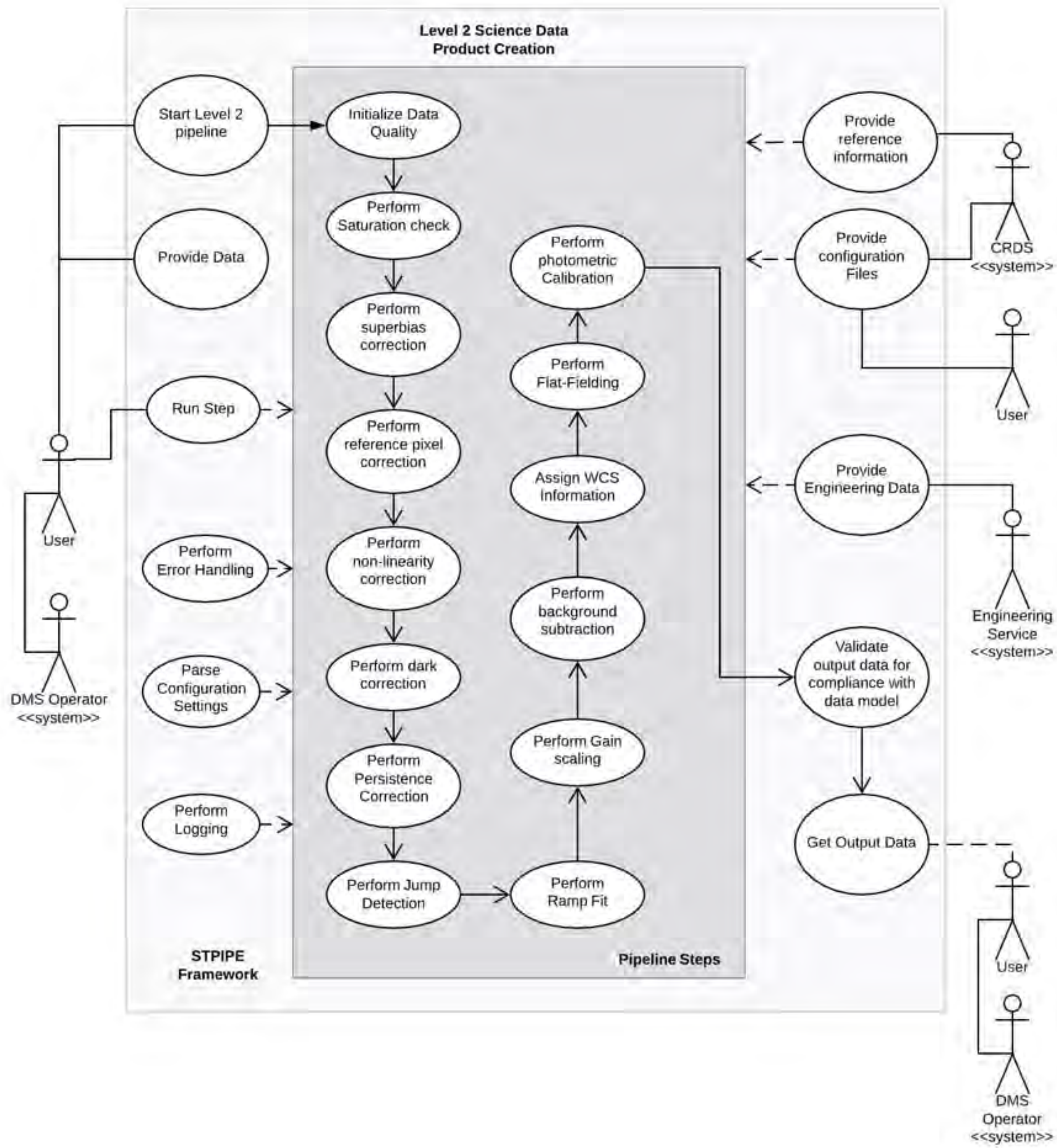
## Data storage & processing

- Cloud-based high-level data processing brings software to the data
- JupyterLab environments ease access, sharing and repeatability
- Software environment for the community in sync with mission data processing

# Roman Data Management



# Exposure Level Processing Flow

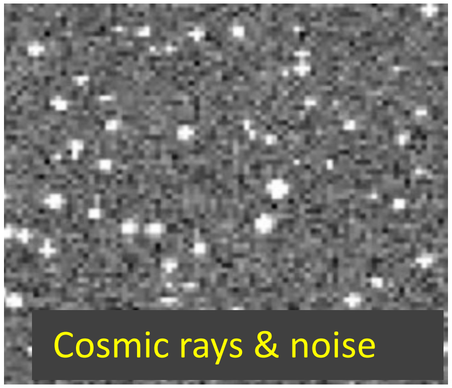


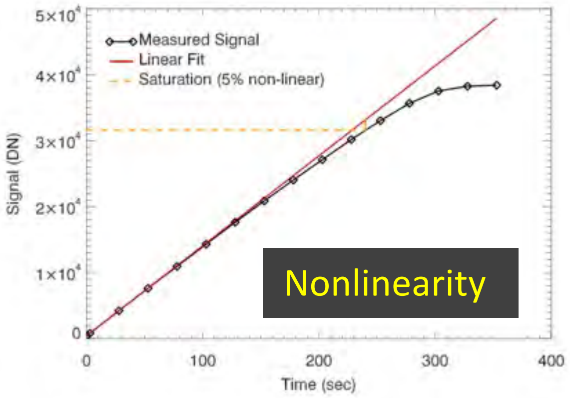
# Data Products

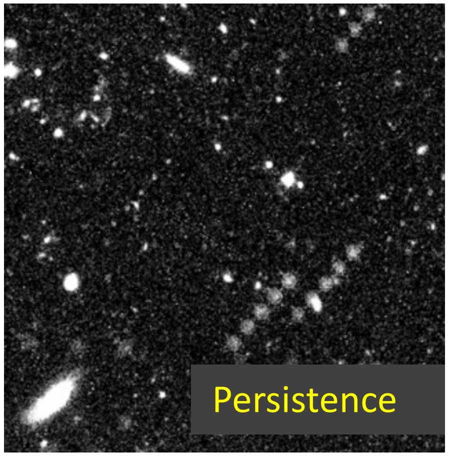
- **Array data**
  - Level 1 – raw
  - Level 2 – instrument signatures removed; aligned to Gaia to <1.3 mas precision
  - Level 3 – rectified and co-added
  - Level 4 – segmentation maps associated with catalogs
  - Queryable Empirical PSF library
  - Level 5 – community contributed products
- **Tabular data**
  - Level 4
    - Static Object catalogs
    - Variability catalogs (from aperture photometry and difference imaging)
    - Idealized source-injection: Catalog of inputs vs. outputs (photometry, sizes, shapes)
  - Level 5
    - Community-contributed products
- **Availability (levels 1-4)**
  - At the individual FOV level within 2 days of receipt of last relevant data
  - Consistently-calibrated data of survey areas released within 6 months

# Simulations: Instrumental Signatures

- **Intended as a tool for pipeline development & testing**
  - Not optimized for use in a pipeline
- **Input**
  - An image in counts/second in the band (produced by idealized simulation tool)
- **Operations**
  - Simulates the individual readouts & the grouping of readouts done onboard
  - Instrument signatures to be included:
    - Cosmic Rays
    - Nonlinearity
    - Interpixel capacitance (if not using an empirical PSF)
    - Basic persistence model
    - Readout noise  & dark current (Simple model)
  - Not included:
    - Intra-pixel sensitivity variations
    - Brighter-fatter effect
    - Sophisticated readout model (e.g. 1/f bias drifts)
- **Output**
  - Image in the format of the inputs to pipeline level 2



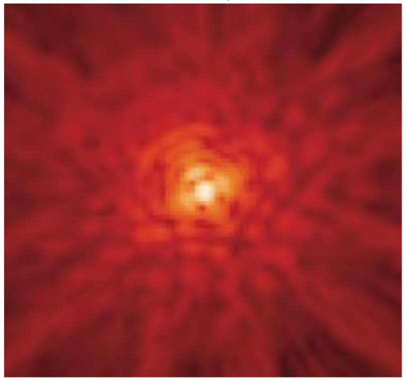
# Simulations: Idealized

- **Intended both to be both by the SOC and as a user tool for custom analysis**
- **Can be used to inject sources into images, or create entirely new images**

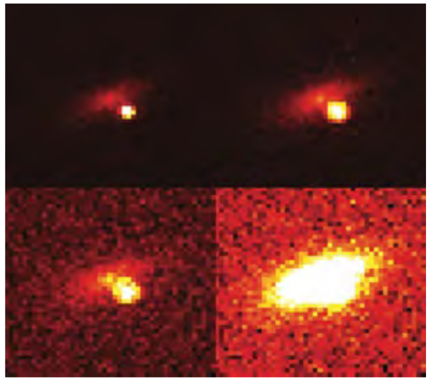
- **Inputs:**
  - Parameters to describe the properties of individual sources
  - Parameters to describe the collection of individual sources
- **Operations:**
  - Apply the instrumental throughputs (using synphot)
    - SEDs or data cube in physical units -> Images in counts/s
  - Convert the model shapes into an image
  - Convolve with the point-spread function
  - Add background
  - Apply geometric distortions (if simulating level-2 outputs)
  - Resample to the appropriate detector pixel geometry
  - Add noise
- **Outputs:**
  - Source "truth" catalog constructed from the input distributions
  - Images in the format of Level 2 output products (i.e. calibrated, but unrectified images), or Level 3 output products (rectified)

# Currently Available Simulation Tools & Applications

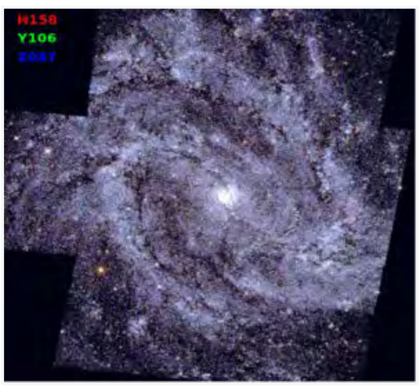
(https://www.stsci.edu/roman/science-planning-toolbox)



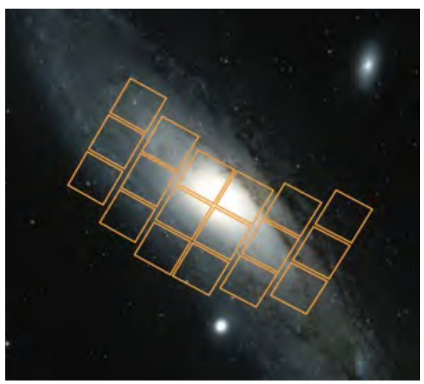
WebbPSF
Wavelength Dependent PSF Simulator



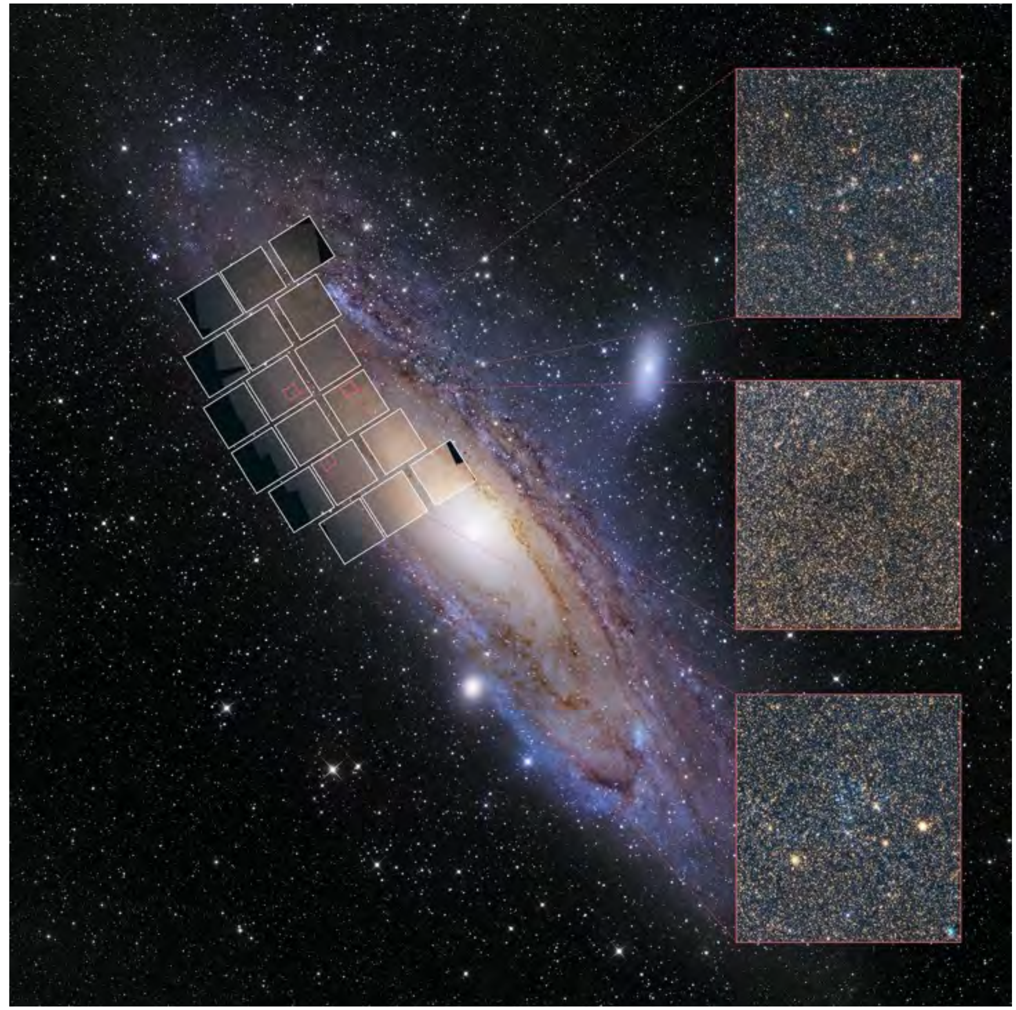
Pandeia
3-D (x,y,$\lambda$) Exposure Time Calculator and Image simulator




STIPS
Image Simulator



Field of View Overlay

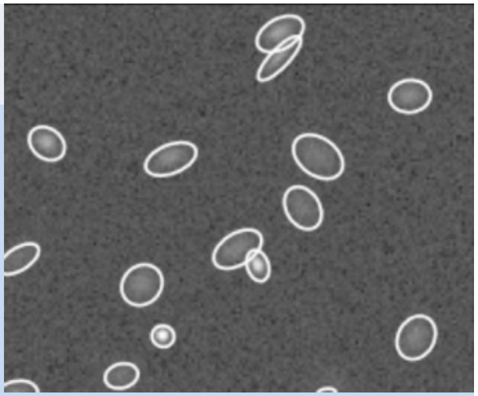


Simulated Roman Observation of Andromeda
(B. Williams with aid of STIPS)

Complementary to tools developed by the community and other partners

# Catalogs: Static

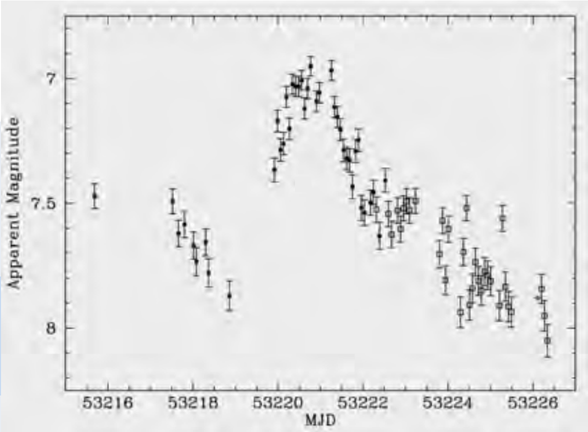
- **Primary cosmology-related goal is to enable accurate photometric redshifts**
- **Accurate, well-characterized photometry and shape measurements will also enable a wide range of general astrophysics**
- **Inputs:**
    - Level 3 images



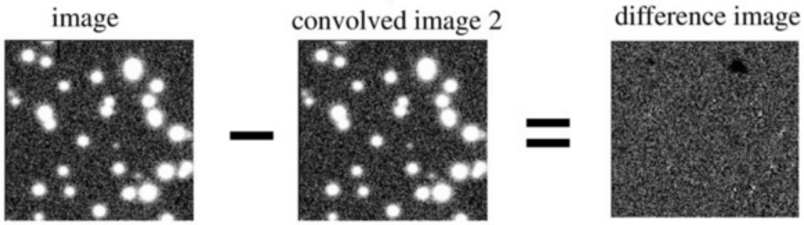
- **Operations**
    - Estimate and subtract background
    - Convolve with a detection kernel
    - Identify connected pixels above a noise-dependent threshold
    - Hierarchically de-blend overlapping sources
    - Measure fluxes through apertures
        - With and without convolution by a PSF-matching kernel to correct all photometry to one reference PSF
    - Measure shapes
    - Star-galaxy classification based on shapes (only)
    - Compute photometric redshifts from multi-band (Roman only) photometry
- **Outputs**
    - Level 4 catalogs including uncertainties computed from noise model
        - Photometry, positions, shapes, & local background estimates
        - Astrometry aligned to Gaia
    - Level 4 segmentation maps
    - Catalogs of input & output parameters for injected artificial sources

# Catalogs: Variability

- **To be run in the pipeline to satisfy the requirement for time-domain information for variable sources**
- **Photometry-based variability catalogs**
  - Inputs
    - Release-level merged survey catalogs
    - Individual images (level 2 or level 3)
  - Operations
    - Compare flux in static catalog to fluxes in individual images
  - Outputs
    - Database of the individual-image photometry for the entire survey area with variability index

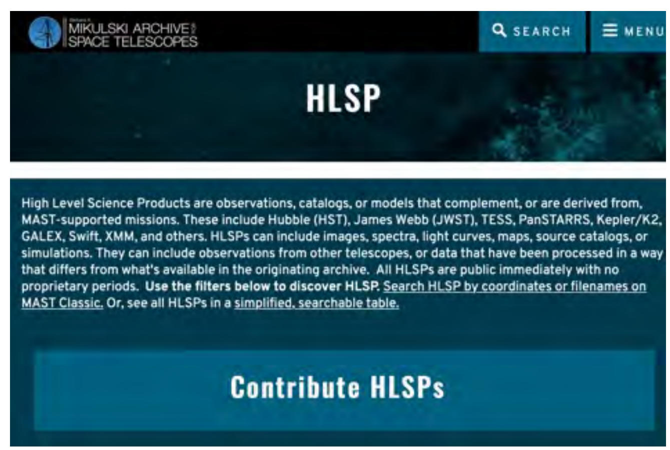



- **Difference-imaging variability catalogs**
  - Inputs
    - Individual rectified images that overlap spatially
  - Operations
    - Convolve with PSF-matching kernel if needed
    - Subtract a template constructed from all but the most recent image
    - Identify point-sources in the difference above a threshold
  - Outputs
    - Level 4 catalog of sources that exceeded the threshold along with associated metadata



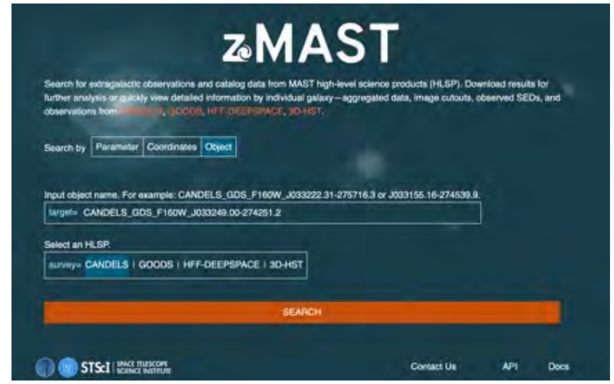
# Community-Contributed Products

- **Public data products contributed by the science community are likely to be widely used. Examples include:**
  - Joint photometry with complementary data sets
  - Photometric redshifts that use complementary data sets
  - Value-added catalogs of derived properties (e.g. from SED fitting)
  - Hybrid spectroscopic and photometric catalogs
  - Survey-level calibrations
    - Improved astrometry & photometry after constraining for consistency across the full survey
    - Window functions, masks, PSF kernels, etc.
  - Transient-free template images
- **Details & cadence to be defined through future community engagement and opportunities**

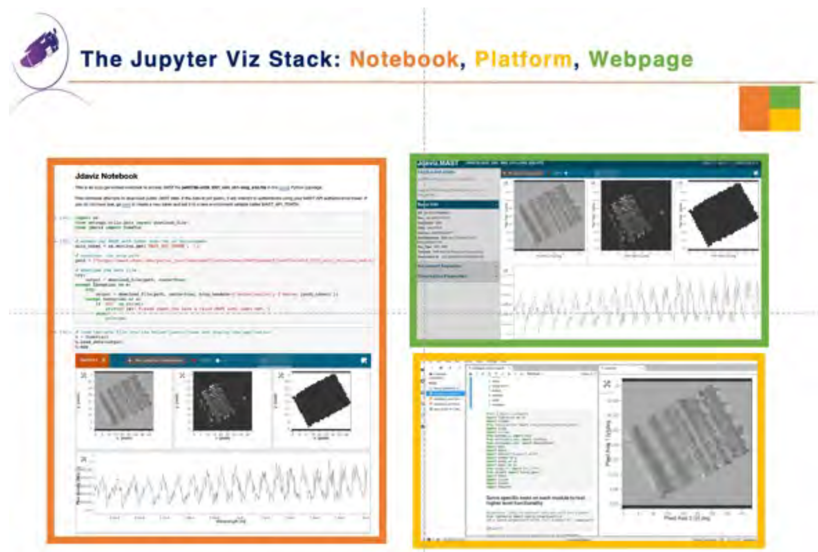


# Archive

- **All Roman Science Data are Public and retrievable from MAST**
  - Includes data from SOC, SSC, Integration & Test, & high-level community products
- **Expect archive services to evolve**
  - Currently incorporating Jupyter analysis & visualization tools for JWST
  - Improving access to high-level products with services like z.mast and exo.mast
- **Higher-level products will be available in the cloud as well**
  - SOC produced Level 2 and beyond
  - SSC high-level products

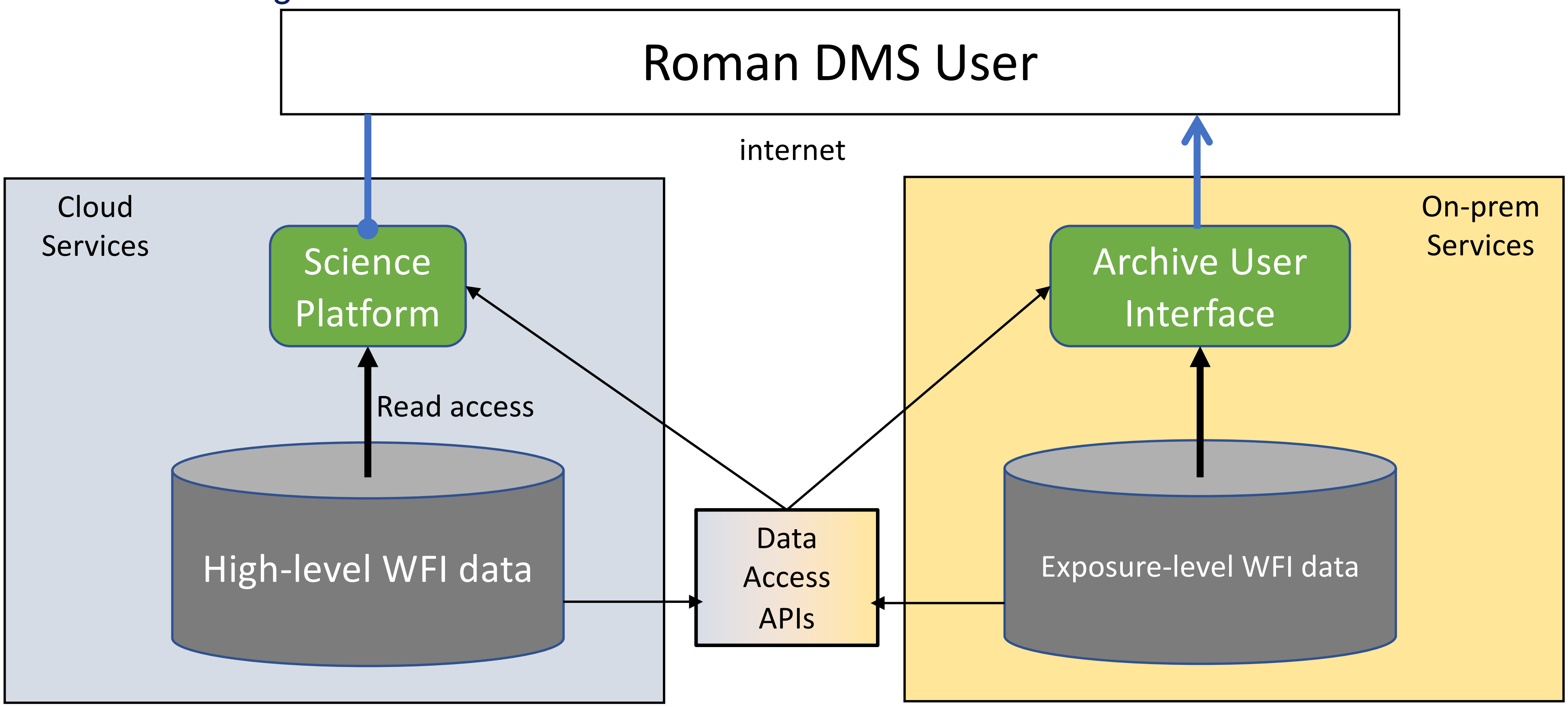






# Accessing Roman Data

- **STScI archive services**
  - Bring data to the users
- **Roman Science Platform**
  - Brings users to the data

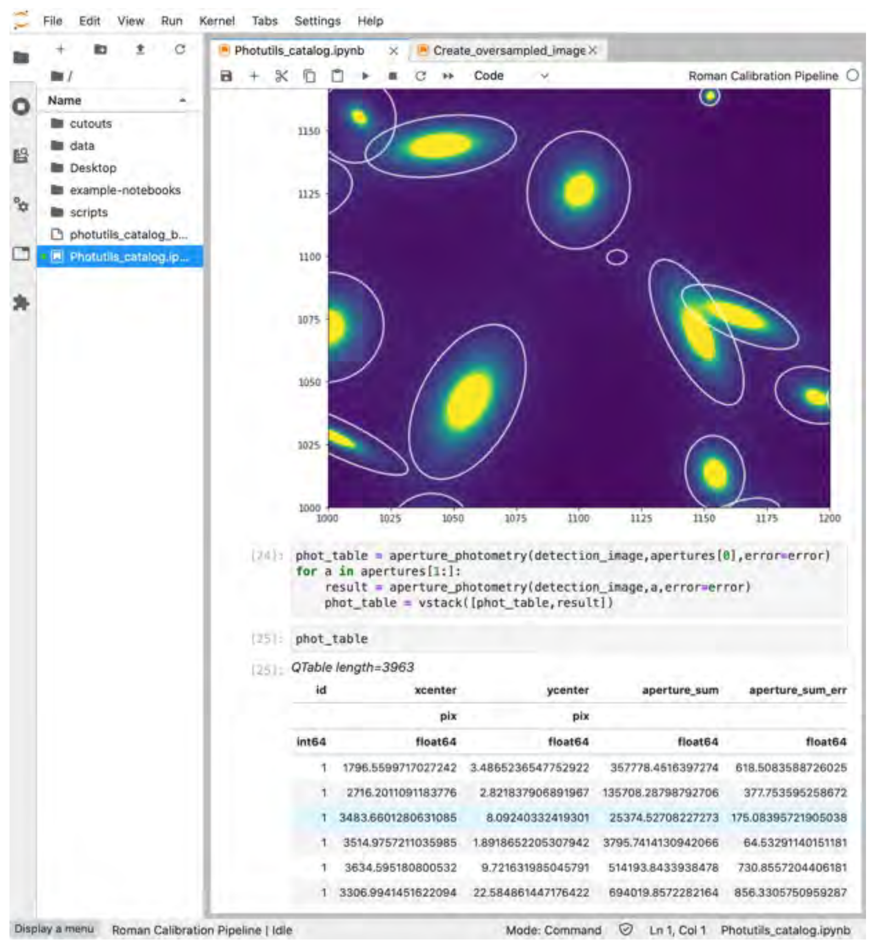


# Why the Cloud?

- **Putting both the computing and the science-ready data in the commercial cloud offers the following benefits:**
  - Convenient scalability for both data volume and computational demands
  - Flexible solutions for specific computing needs (e.g. GPUs or I/O optimized computing)
  - Lower total costs to NASA relative to multiple on-site installations
- **Benefits to the science users include:**
  - Efficient access to the data
  - Computing resources for exploratory work are available with no need to write a grant proposal
  - Local IT and software support costs are greatly diminished
  - Easier collaboration with astronomers across institutions
  - A powerful and stable science software environment

# Community use of the HLPP

- **Log in with your MyST account**
- **JupyterHub instance**
  - Roman science calibration pipeline software installed and configured
  - Full Python + Astropy ecosystem installed and configured
  - Ability to install other packages and your own code
- **Flexible, scalable architecture**
  - Simple to add CPU & storage
  - High-throughput access to the data
  - Can scale up resources (e.g. GPUs or neural engines) as science needs & technology evolves



# Concept for managing Roman platform

- **The platform should provide a sufficient base level of resources**
  - Most NASA Roman grants wouldn’t need a computing line item.
  - Projects needing exceptional resources (“consortia”) could still apply for funds
  - Allows much more global optimization for science than case where funds are locked in small grants.
- **Tier concept to support most users**

| Tier | Users | Typical sky area processed per user |
|---|---|---|
| Entry | 5000 | A few |
| Research | 1500 | Tens to hundreds |
| Team | 50* | Hundreds to thousands |
| Consortium | A few | Thousands |

  - *Team & Consortium: multiple users share resources
  - Will require periodic renewal
- **Looking to to enable ~1000 Roman papers/year**
  - The Roman-data-intensive work; not all the computing
  - Not long-term archival storage of intermediate projects

# Concluding Remarks

- Roman is a unique observatory that will advance a wide range of astrophysical questions, and is complementary to other exciting missions and projects for the coming decade
- In its role as the Science Operations Center, STScI is working collaboratively with all other mission partners and the astronomical community to make Roman a success
- Excellent progress has been made on the design of the Roman SOC; system development has now started
- The Roman Data Management System at STScI will be the key to unlocking Roman's scientific potential, including innovative approaches to address Roman's key Big Data challenges